ARTICLES

THE DOMAIN OF PREFERENCE

LEWIS A. KORNHAUSER

INTRODUCTION

The concept of preference is one of the central, perhaps the central, concept in microeconomic theory. It is also one of the most protean of concepts within economics where it assumes many forms. Its centrality derives from its ubiquity; almost every model in microeconomics specifies a preference ordering for each agent. This very ubiquity, however, explains its protean nature. The models in which these preferences appear have a multitude of purposes. Moreover, a given application for a particular purpose may deploy a very different interpretation or specification of preference than another application that pursues an identical purpose.

Recently, the concept of preference has come under attack in a variety of ways. In this Article, I focus on the most sustained and central attack on the concept—that developed by psychologists and behavioral economists. In Parts I and II of this Article, I briefly characterize the concept of preference itself and the behavioral critique of this concept. I then partially evaluate this critique through two related inquiries.

First, in Part III, I seek a clearer understanding of the multiple uses of the concept within microeconomic theory. These uses fall into four categories: descriptive, explanatory, design, and normative. The

---

1 Alfred and Gail Engelberg Professor of Law, New York University. Mark Geistfeld and Liam Murphy commented on earlier drafts. I also benefited from comments from participants at the University of Pennsylvania Law School Symposium on Preferences and Rational Choice and the University of Chicago Law and Economics Workshop.

behavioral critique of preference theory has different force against different uses across and within categories.

Second, I suggest two reasons why a critique of a particular use of preference apparently undermines the concept of preference in all its uses. In Part IV, I argue that economists have not succeeded in draining their interpretations of the concept of preference of its original psychological content. The critique of behavioral economics undermines the use of preferences in a psychological explanation of behavior. It does not, however, undermine a role for preference within a nonpsychological theory of explanation, or in its design and normative uses. Further, in Part V, I argue that the paradigmatic arena of application of microeconomics—markets for goods and services—is special, and its characteristics do not translate readily to other arenas of application. In the standard model, agents choose from some set of (feasible) options that constitutes a subset of the options over which the agents have (explanatory) preferences. In addition, the well-being of agents derives from the extent to which these explanatory preferences are satisfied. The structure of the standard model is thus special because explanatory, normatively rational, and evaluative preferences coincide. Some models sever the link between the domain of choice and the domain of explanatory preference or between the domain of explanatory preference and the domain of well-being. Once one understands that these domains are not identical, intransitive choices may no longer undermine explanations that rely on a psychological interpretation of the concept of preference.

I. THE FORMAL CONCEPT OF PREFERENCE AND ITS BASIC INTERPRETATIONS

In the discussion of preference, one must distinguish between the formal concept and its interpretations. The formal concept consists of a simple mathematical structure that I discuss below. This mathematical structure, however, may be (and has been) interpreted in many different ways. In a particular application, an interpretation will define very concretely the content of the agent's preferences. These concrete interpretations, however, pass through one of three more general models—decision making under certainty, decision making under risk, and decision making under uncertainty—each of which provides a more abstract interpretation of the formal theory.

The formal concept of preference is quite simple: A preference is a pair \((D, R)\) where \(D\) is a specified domain and \(R\) is a linear order over \(D\). The domain \(D\) is simply a set of "objects." The linear order \(R\)
is a binary relation over the elements of $D$ that satisfies three conditions: (i) it is complete; (ii) it is symmetric; and (iii) it is transitive.\(^2\) Completeness requires that, for every $x$ and $y$ in $D$, either $xRy$ ("$x$ is at least as preferred as $y$") or $yRx$ ("$y$ is at least as preferred as $x$"). Symmetry requires that $xRx$ for every $x$. Transitivity requires that if $xRy$ and $yRz$, then $xRz$; in words, if $x$ is at least as preferred as $y$ and $y$ is at least as preferred as $z$, then $x$ is at least as preferred as $z$.\(^3\)

One may represent many everyday relations, both physical and social, with this formal concept. For example, let the set of objects be mountains on earth and the relation be "at least as tall as." Then this pair $(D = \text{mountains on earth}, R = \text{"at least as tall as"})$ satisfies the requirements of a preference. Or let the set of objects be individuals in the United States and the relation be "at least as old as." This pair, too, satisfies the requirements of a preference. These "preferences," of course, have no economic content.

The three distinct contexts of decision making in which economists deploy the formal concept of preference all use the identical formal concept, even though the terminology used in these contexts—decision making under certainty, decision making under risk, and decision making under uncertainty—differs in order to reflect the different interpretation of the formalism required by the context of application. Under certainty, the decision maker knows with certainty what outcomes will result from her actions. Preferences in this instance are over the domain of "consequences." Under risk, she knows the probability distribution over possible outcomes that is associated with each possible action she may take; economists usually describe the elements of the domain $D$ of preference as "lotteries." Under uncertainty, she knows only the set of consequences that a given action

\(^2\) One should note, at least in passing, that many central results in microeconomic theory do not require preference theory. This role of preference in general-equilibrium results is limited in two respects. First, the existence of general equilibria turns on the excess demand function; it has been shown that one can construct any excess demand function from the excess demand functions of a finite number of consumers. See Gerard Debreu, *Excess Demand Functions*, 1 J. MATHEMATICAL ECON. 15 (1974) (demonstrating such a construction). Second, a general equilibrium will exist even if individuals have acyclic, or even intransitive, preferences. For an example, see theorem 7.21 and the subsequent example in BRYAN ELICKSON, COMPETITIVE EQUILIBRIUM: THEORY AND APPLICATIONS 315-21 (1993).

\(^3\) One may define from the order $R$ two other relations on $D$: the relation $I$, which I call "indifference," and the relation $P$, which I call "strict preference." \(I\) is defined by the requirement $xIy$ if and only if $xRy$ and $yRx$. Indifference is a partial order. The definition of strict preference states that $xPy$ if and only if $xRy$ and not $yRx$. 

may have; economists usually describe the elements of the domain $D$ of preference here as "actions."

Even the simple characterization of the domain of preference as "consequences," "lotteries," or "actions" begins the process of interpretation. Formally, the set $D$ in models of decision making under certainty is simply a set of objects; in the standard market context, the objects are "consumption bundles." In decision making under risk, the domain of preference is formally the unit simplex $\Delta$ of some (finite) set $s$. In decision making under uncertainty, the domain $D$ of preferences is a set of functions from a (finite) set $s$ (generally interpreted as "states of the world") to a set $X$ (generally interpreted as "consequences"). In each context, microeconomic theory then generally proceeds to identify a set of conditions that are necessary and sufficient for the representation of the preference by a "utility" function or by "expected utility."\(^5\)

II. THE BEHAVIORAL CRITIQUE OF PSYCHOLOGICAL PREFERENCES

Research in behavioral economics has demonstrated a number of ways in which individual decisions deviate from the dictates of rationality as understood in preference theories. I consider a small set of these results that bear directly on preference theory.\(^6\)

First, the initial allocation of property rights affects an individual's valuation of a good; an individual is willing to pay less to acquire a good than she is willing to accept to sell it. This endowment effect ex-

---

4 The unit simplex in $\mathbb{R}^n$ consists of all points $s=(s_1, s_2, \ldots, s_n)$ with $s_i \geq 0$ for all $i$ and such that $\sum s_i = 1$.

5 For typical representation theorems, see Andreu Mas-Colell et al., Microeconomic Theory 8-9 (1995).

6 Other evidence suggests a number of individual deviations from rational decision making as articulated in preference theories. First, individuals in certain circumstances do not conform to the norms of logic. See Leda Cosmides & John Tooby, Cognitive Adaptations for Social Exchange, in The Adapted Mind 163, 167-70 (Jerome H. Barkow et al. eds., 1992) (discussing altruistic motivations of behavior). Second, individuals do not act in a purely self-interested fashion. See, e.g., Alvin E. Roth, Bargaining Experiments, in The Handbook of Experimental Economics 253, 296-302 (John H. Kagel & Alvin E. Roth eds., 1995) (discussing the ultimatum gaming experiment). Finally, norms of fairness seem to govern some individuals' actions. See, e.g., Ernst Fehr & Klaus M. Schmidt, A Theory of Fairness, Competition, and Cooperation, 114 Q.J. ECON. 817, 818 (1999) (finding that "in addition to purely self-interested people, there are a fraction of people who are also motivated by fairness considerations"); Ernst Fehr & Simon Gächter, Fairness and Retaliation: The Economics of Reciprocity, J. ECON. PERSP., Summer 2000, at 159, 159 (suggesting that "many people deviate from purely self-interested behavior" in certain social contexts).
ists even when the wealth effect of a particular allocation is small. Kahneman, Knetsch, and Thaler, for example, found an endowment effect when the market value of a fungible good was roughly three dollars. The endowment effect permits intransitive preferences. Consider some good X for which the agent has a willingness to accept of $510 and a willingness to pay of $490. Let bundle A consist of X and $500, bundle B of $1000, and bundle C of X + $500. Then if the agent has bundle A, she prefers A to B; if she does not own X but has $1000, she prefers B to C, but she should be indifferent between A and C. These pair-wise comparisons yield the judgments A preferred to B preferred to C which is as preferred as A. A violation of the transitivity requirement undermines each of the three interpretations of preference theory as decision making under certainty, risk, and uncertainty.

The endowment effect leads to intransitivity and a violation of the assumptions of preference theory only under some descriptions of the domain over which the agent has preferences. The violation arises when we describe options within the domain solely in terms of the physical characteristics, date, and location of the goods at issue. If we supplement this physical description of the bundles with the ownership status of the goods, then the intransitivity disappears. The agent prefers A to B only if she owns; otherwise she prefers B to A. Ownership status is thus decision relevant.

This strategy to rescue preference theory is unsatisfying. Such a redescription is always available but not every redescription is plausible. In the absence of a reason why a particular factor should influence the agent’s valuations, the redescription seems merely ad hoc. To require a reason to justify treating a particular attribute of a situation as decision relevant suggests that some preferences are irrational. It thus apparently imposes substantive content on the idea of a preference. Appearances may deceive; the endowment effect poses a problem because agents systematically exhibit this behavior. It thus seems

---

7 See Daniel Kahneman et al., Experimental Tests of the Endowment Effect and the Coase Theorem, 98 J. POL. ECON. 1325, 1346 (1990) (concluding that endowment effects are “fundamental characteristics of preferences”). The low price of the good eliminates the possibility that the allocation of the property right significantly affects an individual’s wealth; even an individual whose preferences depended on wealth would have identical preferences regardless of the allocation of the right. Fungibility eliminates any possibility that a gap between offer and asking price would arise from nonsubstitutability.
important to explain either causally or normatively why they exhibit this behavior.ª

Second, individuals do not choose under risk in a fashion consonant with the dictates of preference theory.º Their probability judgments are defective. For example, individuals generally exhibit hindsight bias; they overestimate the likelihood that what actually happened would happen. This bias may arise in any legal context that requires a fact-finder to assess ex post the reasonableness of ex ante conduct that led to undesirable results. Hindsight bias is, arguably, an instance of a more general bias that individuals systematically exhibit. Agents overestimate the probability of "salient" events. So, for example, individuals do not accurately assess information about unlikely events, such as disasters or crimes, that are prominently reported in the news media. Thus, individuals may believe very low probability events to be more likely to occur than they in fact are.

Third, and most important, some research suggests that individuals do not have well-defined preferences at all. The expressed preference depends on the way in which the preference is elicited. Consider, for example, the phenomenon known to economists as preference reversal, which was first studied in 1971.¹⁰ These experiments elicit preferences for options in two different ways. A subject is asked to choose directly between options \(H\) and \(L\) and she is asked to price the two options (by assigning prices \(p_H\) and \(p_L\) respectively to the \(H\) and \(L\) options). If the individual has well-defined preferences, she should choose \(H\) over \(L\) if and only if \(p_H\) is greater than \(p_L\). Consider for example the following two options. First, let \(H\) be the lottery: win $10 with probability \(7/9\) and win nothing otherwise. Second, let \(L\) be the lottery: win $100 with probability \(1/12\) and win nothing otherwise. Studies show that individuals generally choose the lottery \(H\) over the lottery \(L\), but that they generally assign a (lowest selling) price to \(L\).
that exceeds the (lowest selling) price they assign to $H$. This assignment of prices suggests that the individuals generally prefer $L$ to $H$.

One might explain the phenomenon of preference reversal in three different ways. First, preference reversal might result because individual preferences are simply not transitive. Second, because the choices involved concern lotteries, preference reversals might result because individual preferences over risky outcomes do not conform to the independence axiom. Finally, preference reversals might result because individuals do not have well-defined preferences. Using a clever experimental design, Tversky, Slovic, and Kahneman produced results suggesting that a substantial part of the phenomenon is best explained by the third of these causes: individuals do not have well-defined preferences. Other studies show a similar dependence of preference on the method of elicitation. Generally, individuals respond differently to risky choices that are phrased as providing gains than to risky choices that are phrased as providing losses.

More recent empirical work suggests that the phenomenon of preference reversal presents less of a threat to the psychological explanatory project of preference theory. One study, rather than using questionnaires, used real payments to the subjects. It found significantly lower rates of preference reversal when subjects received real payments. A second study argued that individuals would perform better if presented with choices more akin to those presented in real markets. In this experimental test, the preference reversal rate fell to eleven percent, significantly lower than that observed in other experiments. A third study showed that the rate of preference reversals

---


12 See Barbara J. McNeil et al., On the Elicitation of Preferences for Alternative Therapies, 306 NEW ENG. J. MED. 1259, 1261-62 (1982) (analyzing the results of an experiment in which subjects were asked to choose between forms of medical therapy), cited in Amos Tversky & Daniel Kahneman, Rational Choice and the Framing of Decisions, in ROBIN M. HOGARTH & MELVIN W. REDER, RATIONAL CHOICE: THE CONTRAST BETWEEN ECONOMICS AND PSYCHOLOGY 67 (1987). One might offer an alternative explanation of the McNeil study that does not undermine the concept of preference; rather, the subjects may reveal their inability properly to evaluate probability distributions. The different phrasings of the two treatments may highlight different portions of the probability distribution.


14 Id. at 1377.


16 Id. at 341.
dropped more dramatically when subjects acted repeatedly in a market setting than when an individual is confronted with an isolated set of choices.\textsuperscript{17} If the agent's expression of preference depends on the procedure used to elicit that expression, we may conclude that the agent does not possess a stable preference that guides her choices. If this phenomenon were widespread, it would render preference theory unusable as a psychological theory of explanation.

III. THE PURPOSES OF PREFERENCE

Economic models serve several different functions. One may identify at least four distinct categories of functions. Models may serve: (1) to summarize actual choices; (2) to explain choices; (3) to facilitate the design of institutions; and (4) to evaluate choices. Within each of these categories, one might develop models for radically different purposes. Each purpose prompts a different interpretation of the formal concept of preference.

A. Summarizing Choice

This approach takes the individual's choices as primitive. An agent is presumed to have a choice function \( C(s) \) which identifies, from each possible choice set \( s \), the element (or elements) that the agent chooses.\textsuperscript{18} The theory then identifies the conditions under which the choice function \( C(s) \) can be represented by a preference relation \( (D, R) \).\textsuperscript{19} On the revealed preference account, the preference does not explain the choices; it merely summarizes them in a convenient way. Revealed preference theory thus implicitly assumes that the agent has preferences over the domain from which she chooses. One might say that the domain of preference is assumed identical to the


\textsuperscript{18} Technically, one must identify the domain \( P(D) \) of the choice function \( C(-) \) where \( D \) is a set of elements and \( P(D) \) is the power set of \( D \)—i.e., the set of all subsets of \( D \).

\textsuperscript{19} Contraction consistency (condition \( a \)) and expansion consistency (condition \( b \)) are necessary and sufficient conditions for the choice function to be a preference. One may illustrate these conditions in the context of chess players. Informally, contraction consistency requires that, if the world chess champion is a Russian, then the world chess champion is also the Russian chess champion. Similarly, expansion consistency requires that, if the world champion is a Russian, then all Russian chess champions are world chess champions.
domain of choice. I pursue the significance of the elision of these two domains in Part V below.

Revealed preference theory arose in the context of the theory of the consumer. This theory "naturally" specified the domain of choice as the set of possible "consumption bundles" understood as the basket of commodities produced in the economy. Under this interpretation, it is reasonable to assume that the agents have preferences directly over the set of consumption bundles from which they choose. Nevertheless, an agent's choice function might not satisfy the necessary and sufficient conditions for the existence of a preference that represents that choice function. Thus, revealed preference theory is not tautological.

B. Explaining Choice

Many economic applications seek to explain, rather than to summarize, the choices of individuals. One might adopt at least two distinct explanatory strategies: a psychological one or a functional one. The psychological strategy interprets the individual's preference psychologically—that is, as some mental entity or some proposition towards which the agent holds an attitude of some sort. On this account, preference theories explain choices in a causal way: the agent chooses a given feasible option because she desires it most or believes that, all things considered, it best promotes her ends.20

Much of the literature critical of preferences as an explanatory model assumes explanation runs through the psychology of the agent. Amartya Sen's early and influential critique of microeconomics essentially argues that preference theories are only amenable to impoverished interpretations of individual psychology.21 Similarly, a psychological interpretation of preference theory underlies Alexander Rosenberg's critique of microeconomic theory—he argues that the theory rests on a formalization of folk psychology.22 Folk psychology,

20 I ignore here philosophical problems concerning the causal power of reasons for action.
22 ALEXANDER ROSENBERG, MICROECONOMIC LAWS: A PHILOSOPHICAL ANALYSIS (1976).
according to Rosenberg, is fatally flawed as a basis for empirical science and offers no room for improvement.25

The behavioral-economics challenge to the neoclassical paradigm of preference maximization presumes that preference theory provides a psychological explanation for behavior. The empirical research consists of experiments conducted in the laboratory and hence motivated by nonpsychological factors. The research thus asserts, essentially, that individuals do not choose in accordance with expected utility theory. Another interpretation consistent with the psychological strategy regards a preference as a summary of the agent's all-things-considered judgments of what she ought to do. For this interpretation to be explanatory, the investigator must assume that agents (generally) do what they ought to do. Under this interpretation, an agent may have moral motivations as well as self-interested ones. One might understand this interpretation as a strategy of incorporation because it subsumes all decision-relevant concerns that agent might have into her preference ordering. Of course, it may now become difficult to assure that these concerns constitute a well-defined ordering over the agent's choices.24

The functional strategy avoids interpreting preferences as psychological states. Rather, it explains a choice Y as promoting some end Z whether or not the agent consciously pursues that end.25 Functional explanations, of course, are controversial because they appear to explain a cause by its effect—in this case, the choice Y is explained in terms of its effect on Z. A more satisfying explanation would identify some causal mechanism that explained how choice Y came to promote end Z.26 The functional strategy permits the attribution of a

---

23 Id.


25 For an endorsement of this strategy of explanation, see Debra Satz & John Ferejohn, Rational Choice and Social Theory, 91 J. PHIL. 71 (1994).

26 For a discussion of functional explanation, see Jon Elster, The Cement of Society (1989). See also Jon Elster, Marxism, Functionalism, and Game Theory: The Case for Methodological Individualism, in Marxist Theory 48, 49-55 (Alex Callinicos ed., 1989) (comparing the historical role of functional explanation in several disciplines against its role in the Marxist theory); G.A. Cohen, Reply to Elster on 'Marxism, Functionalism, and Game Theory,' in Marxist Theory, supra, at 88, 104 (claiming that "functional explanation lies at the heart of historical materialism" and that "game theory therefore cannot replace functional explanation within Marxist social analysis").
preference to "agents" that are not conscious. One may thus speak, for example, of the "selfish gene" that seeks to maximize its "preference" defined in terms of reproductive fitness. This attribution is particularly compelling because, in the biological setting, natural selection provides a mechanism that ensures that the agent's "end," $Z$, is in fact pursued by the agent.

In economics, a similar selection mechanism is also often invoked. Market settings, it is argued, select for profit maximizers. Regardless of the individual aims of market participants, then, one will observe behavior explained by the imputed set of preferences. Whatever one's view of the success of the functional strategy of explanation in market settings, it is more problematic in most legal and political applications. Neither legal nor political settings present a competitive environment that plausibly selects for agents that maximize the appropriate objective function.

C. Design

Much research in economics and economic analysis of law addresses policy concerns rather than explanatory ones. Economic theory in general, and the concept of preference in particular, plays two distinct roles in policy analysis. First, economists often invoke a preference-based normative criterion against which to assess policy options. I address this role of the concept of preference in the next section. Second, because virtually every criterion for the evaluation of policy is consequentialist, policy analysis must predict the consequences of the various policy options. Economic theory provides one such predictive theory.

A good predictive mechanism need not be an explanatory theory. Reliable prediction requires only that one identify features of the world that are well-correlated with the outcomes that one wishes to predict. These correlated factors need not be causally related and hence may not explain. Much policy prediction in fact does not rely on explanatory theories. When Congress seeks to predict the revenue consequences of possible changes to the tax code, it does not attempt to predict in detail the responses of individuals to the changes in the

---

27 See Prajit K. Dutta & Roy Radner, Profit Maximization and the Market Selection Hypothesis, 66 REV. ECON. STUD. 769, 770 (1999) (providing a model in which profit-maximizing firms go bankrupt in finite terms with mobility, but noting that there exist other strategies with positive probability of infinite existence).
tax code. Prediction based on proxies may perform better than predictions based on explanatory theories.

D. Normative Uses

The concept of preference plays a central role in two distinct, normative projects. First, preference theory often serves as a normative baseline for instrumental rationality. In this normative role, the content of the agent's preference is irrelevant—the conception of rationality at issue requires only the formal structures of preference to provide a baseline. Knowledge of an individual's preferences coupled with the relevant preference-based decision theory permits one to assess what an individual ought to do to best promote her own aims. This role for preference theory contrasts with the understanding underlying behavioral economics. That theory has no normative aim. Failure to act as behavioral theory predicts does not subject the agent to criticism for acting irrationally; rather, it only gives the theorist reason to revise her theory.

The second normative role for the concept of preference in economic theory is evaluative—it depends not only on the structure of preference but also on the content of the preference of the specific agent. Consequently, the evaluative role of preference requires an interpretation of the domain D. This role for preference evaluates actions, policies, institutions, or states of the world in terms of the effects of the action, policy, or institution on the agent's well-being understood as the extent to which the agent's preferences are satisfied. Consider for example the use of preference to evaluate policies or actions. In such a model, each agent's preference is interpreted as well-being. Well-being itself may be understood variously. Some understand well-being as an objective concept that does not correspond (en-

28 For a discussion of the difficulties of predicting tax revenues and the various nonbehaviorally grounded models used, see George R. Zodrow, Economic Analyses of Capital Gains Taxation: Realization, Revenues, Efficiency and Equity, 48 Tax L. Rev. 419, 430 (1993).

29 For example, short-term forecasts of weather that predict weather at time $t$ simply by looking at the weather at time $t-1$ may perform much better than predictions based on explanatory models. Similarly, one might predict earthquakes more reliably by studying the behavior of animals than by using geological models.

30 One might regard either preference theories or behavioral theories as theories of rational competence. One may then distinguish, as in linguistics, competence from performance. From this perspective, failure to act as the theory requires indicates that the agent's performance fails to match her competence. See Edward Stein, Without Good Reason 37-78 (1996) (discussing the theory of competence in linguistic theory).
to any psychological state. The objective account of well-being may nonetheless satisfy the formal requirements of a preference, but an objective, evaluative account of preference must then differ fundamentally from an explanatory conception of preference. Such a difference will persist even if, following other understandings, we provide a psychological account of well-being—that is, one that refers to the mental states of the agents. These psychological conceptions of well-being, however, need not correspond to a psychological account of the explanatory conception of preference. An individual’s motivations need not correspond to the set of psychological states that determine the agent’s well-being.

IV. PREFERENCE AND PSYCHOLOGICAL EXPLANATION

The term “preference” suggests a psychological interpretation of the formal concept. Indeed, much confusion, both among economic practitioners and critics, stems from the myriad ways in which psychology infuses some understandings of preference theories but not others. This Part suggests several reasons why divorcing preference from psychology has proved difficult. I then argue briefly that the behavioral economics critique of preferences assumes that preference can be explained psychologically. Finally, I consider the analogy to biology to illustrate a conception of preference without psychology.

A. The Difficulties of Divorce

A sketch of the intellectual history of preference theories provides a useful starting point for this discussion. Over the course of the first half of the twentieth century, economic theorists gradually drained the concept of preference of content other than a summary of choices. In early incarnations, the concept of preference had strong links both to conceptions of well-being within the utilitarian tradition and to conceptions of the psychology of human action. Classical utilitarianism, in fact, draws on both traditions as its conception of well-being derives from its conception of human action and motivation. Classical economists, who both preceded and succeeded Jeremy Ben-

---

51 For discussions of the conception of well-being, see T.M. SCANLON, WHAT WE OWE TO EACH OTHER 108-88 (2000); L.W. SUMNER, WELFARE, HAPPINESS AND ETHICS (1996); Kornhauser, Preference, Well-Being and Morality, supra note 24.

than's utilitarian influence, however, often had a more complex conception of human motivation than the one embodied in current conceptions of economic rationality. An important strain of current economic thinking distilled from these earlier conceptions all reference either to motivation or to well-being. This process of distillation yielded an understanding of preference as simply a summary of choice. In 1938, Paul Samuelson formalized the approach in the *weak axiom of revealed preference*. Subsequently, other authors have extended revealed preference theory.

Two important features of current preference theories had origins in the hedonism of classical utilitarianism. Classical utilitarians explained behavior with a theory of psychological hedonism—an individual acted to maximize "utility" with pleasure providing positive utility and pain negative utility. Similarly, the moral theory of classical utilitarianism required the individual to maximize social utility, understood as the sum of the utility of each person in society. Thus the psychological state of pleasure was central to both the explanatory and evaluative theories of classical utilitarians.

Subsequent generations of economists have struggled to free both the explanatory and evaluative projects from their utilitarian origins. With respect to explanatory preferences, economists attempted to distill away all psychological content. Economists had a less extreme aim for the evaluative project. There, they merely strove to sever the link between evaluation and classical utilitarianism.

The work of Vilfredo Pareto signaled a major shift in the foundations of both the explanatory and evaluative projects of economics. Pareto derived consumer demand theory from an ordinal conception of utility; that is, Pareto showed that microeconomic theory required

---

only the structure of a formal preference to explain market behavior.\textsuperscript{37} Pareto also proposed an evaluative criterion that relied only on ordinal utility information. The partial order, "Pareto superior to," and the maxima of this relation, the "Pareto optimal" points, require only ordinal comparisons of well-being.

The detachment of the concept of preference from its utilitarian origins increased dramatically over the next forty years. First, in the context of welfare economics, the view that interpersonal comparisons of well-being were not possible became entrenched.\textsuperscript{38} The widespread acceptance of this view led to efforts to construct evaluative criteria that avoided such comparisons. The Kaldor-Hicks compensation test constitutes one such attempt.\textsuperscript{39} Second, as noted above, revealed preference theory developed further to detach explanatory preferences from psychological roots. As with evaluative preferences, the aim in part derives from the practical difficulties in determining someone's preferences.

These efforts to purge psychological content from explanatory preference theories have failed. The structure of the formal theories of decision making under certainty, risk, and uncertainty provide the most direct evidence of this failure. The centerpiece of these theories, perhaps their aim, is a representation theorem. A representation theorem identifies the set of axioms that an agent's preferences must satisfy in order for her preferences to be represented by a utility function (with an appropriate structure).\textsuperscript{40}

\textsuperscript{37} At least, this conclusion flows from Pareto's argument. If one requires only ordinal utility information, then the utility function serves only as a convenient representation of an agent's underlying preferences.

\textsuperscript{38} "Impossibility" has two different interpretations. One might regard interpersonal comparisons as conceptually impossible or meaningless. Alternatively, interpersonal comparisons might be practically impossible. LIONEL CHARLES ROBBINS, AN ESSAY ON THE NATURE AND SIGNIFICANCE OF ECONOMIC SCIENCE (3d ed. 1984), is the classic text on the impossibility of interpersonal comparisons. One might understand Arrow's inclusion of the axiom, independence of irrelevant alternatives, as flowing from this tradition. INTERPERSONAL COMPARISONS OF WELL-BEING (Jon Elster & John E. Roemer eds., 1991), is a more recent collection of essays that address the problems of interpersonal comparability.


\textsuperscript{40} See generally DAVID M. KREPS, NOTES ON THE THEORY OF CHOICE (1988) (providing an excellent introduction to representation theorems and considering representation of preferences under risk and uncertainty). In chapter 3, Kreps discusses representation of preferences under certainty. Id. at 19-30.
The representation theorem in the theory of decision making under certainty is the simplest and has the least psychological content. It states that any preference that satisfies a "closure" condition can be represented by a continuous utility function. Neither the closure condition nor the characteristic of continuity has any strong psychological content. The representation theorem is useful, however, for reasons of mathematical tractability; standard techniques of optimization may be applied to the utility representation of closed preferences. The representation theorems for decision making under risk and under uncertainty, however, have significantly more psychological content. In each of these theories, the major result is a theorem that identifies the restrictions on preferences necessary to represent the preference by an "expected utility." An expected utility representation characterizes the agent's preferences in terms of "beliefs" and "desires." Both beliefs and desires are concepts drawn from (folk) psychology, and their use infuses the theories with psychological content. The characterization states that there exists a "utility function" on consequences and a set of beliefs such that an agent prefers some lottery or action \( x \) to another lottery or action \( y \) if and only if the expected utility of \( x \) exceeds the expected utility of \( y \). In the theory of decision making under risk, the agent's beliefs are the objective probabilities that each state of the world will occur. In the theory of decision making under uncertainty, the agent's beliefs are subjective probabilities that each state of the world will occur.

---

41 Let the domain of preference be \( D \). The closure condition requires that, for every \( y \) in \( D \), the set of \( x \) that are at least as preferred as \( y \) and the set of \( x \) such that \( y \) is at least as preferred as \( x \) are closed. The closure condition is usually called "continuity." See GERARD DEBREU, THEORY OF VALUE: AN AXIOMATIC ANALYSIS OF ECONOMIC EQUILIBRIUM 55-59 (1959) (discussing "continuous" utility functions).

42 The most mathematically tractable class of preferences can be represented by a twice continuous (concave) utility function. A preference must satisfy additional restrictions to belong to this class. See Gerard Debreu, Smooth Preferences, 40 ECONOMETRICA 603 (1972) (describing the necessary conditions for preference relations of different classes). Not all preferences satisfy either the continuity conditions nor the additional conditions; moreover, some such preferences are actually tractable. Lexical preferences are the most common example. DEBREU, supra note 41, at 72 n.2.

43 For decision making under risk, the agent's preferences over lotteries must not only be continuous, but must also satisfy a reduction and an independence axiom. For definitions and discussion of these axioms, see MAS-COLELL ET AL., supra note 5, at 167-82. For decision making under uncertainty, additional restrictions, such as some interpretation of the sure thing principle, are required. For definitions and discussion, see id. at 204.

44 Id. at 204.
will be realized. The agent’s utility, in both cases, represents the agent’s preferences over riskless (or certain) consequences.

At least two other factors also explain the difficulty in draining the concept of preference of its psychological content. First, a psychological interpretation of explanatory preferences emerges in part from the connections of the explanatory use to the two normative uses of preference theory. Second, some argue that the requirement of transitivity of choices is sensible only under a psychological interpretation of preferences. I consider these factors in turn.

Both normative uses of preference theory as criteria of rationality and as a standard against which to assess institutions and behavior suggest a psychological interpretation of the formalism. Consider first the evaluative use of preferences as well-being. In most economic models, well-being is understood subjectively in terms of the satisfaction of the individual’s preferences. This understanding has a psychological cast. In the standard context of markets, it is natural to elide the agent’s preferences understood as well-being to her “motivational,” and hence explanatory, preferences. Indeed, the first and second welfare theorems in general equilibrium theory do so explicitly. The correspondence between competitive equilibria and Pareto optimality links the explanatory and evaluative programs together.46

Consideration of preference theory as a criterion of rationality leads to a similar conclusion. When preference theory serves as a criterion of rationality, the analyst recommends its structure as the appropriate means of reasoning about means to ends. The theory states that an individual’s decision processes ought to conform to the axioms of the theory. An individual whose preferences do not conform or who lacks preferences altogether has not decided appropriately. Obviously, an agent who conforms her decisions to the normative demands of preference theory will, in fact, behave in conformity with that theory. The theory will both explain and justify her behavior.47

Amartya Sen has argued that revealed preference theory does not succeed in eliminating psychology from the concept of preference.

---

45 On these theorems, see Debreu, supra note 41, at 94-96.
46 Id.
47 Some philosophers have argued that understanding others requires an imputation of rationality to them. See Donald Davidson, Actions, Reasons and Causes, in ESSAYS ON ACTIONS AND EVENTS 3 (1980) (defending the “ancient—and commonsense—position that rationalization is a species of causal explanation”); Daniel C. Dennett, True Believers: The Intentional Strategy and Why It Works, in THE INTENTIONAL STANCE 13, 15 (1987) (advocating an “intentional strategy” that “consists of treating the object whose behavior you want to predict as a rational agent”).
He contends that the force of the transitivity condition on choice derives from its connection to psychology. Though it may be that nothing in the concept (or act) of choice requires transitivity, it is not clear that anything in the concept of preference requires it either. Usually, transitivity is defended with a "money pump" argument. Someone with intransitive preferences (and no foresight) may be induced into a series of transactions that make her worse off, as measured by her own preferences.

B. Preference Without Psychology: The Analogy to Biology

As noted earlier, an explanatory preference theory does not require a psychological interpretation. One might argue that some process of "selection" insures that agents maximize. Evolutionary biology adopts such a hypothesis to explain the adaptation of species to their environment. Similar selective explanations have been offered in economics and economic analysis of law.

Identification of a selection mechanism will usually require the specification of the content of the individual agent's preferences. In the standard market context of economics, agents are generally assumed to have self-interested preferences. Consumers care only about their own consumption, and firms seek to maximize profits. The claim that the selection pressure created by competition induces agents to maximize preferences has the most plausibility in this context. It appears, however, that, even in this context, the claim is suspect. The threat of bankruptcy is not sufficient to insure that only profit-maximizing firms survive. Of course, profit maximization might be assured through another selection mechanism; perhaps shareholders who seek to maximize wealth select for managers who maximize profits.

The environments in which individual agents in economic analysis of law act generally seem less likely to select for individuals who maximize preferences. In the late seventies, Paul Rubin and George Priest suggested that common law rules would be efficient because in-

---

48 See Amartya Sen, Behaviour and the Concept of Preference, 40 ECONOMICA (n.s.) 241, 242-44 (1973) (claiming that all contributions to revealed preference theory must either implicitly or explicitly involve psychological assumptions).


50 See Dutta & Radner, supra note 27, at 770 (positing that, regardless of threatening consequences, "after a long time, most... surviving firms will not, in fact, be profit maximizers" (emphasis omitted)).
efficient rules would be litigated more often than efficient ones. Subsequent work has challenged this claim on two grounds. First, it is not clear that inefficient rules will in fact be litigated more often than efficient ones. Second, even if this differential litigation hypothesis is true, the selection argument may not in fact apply.

V. THE DOMAIN OF PREFERENCE

A. The Coincidence of Uses

Many models in both economics and economic analysis of law serve multiple purposes. These models assume implicitly that explanatory and evaluative preferences are identical. In some cases, they may assume that descriptive preferences also correspond to the explanatory and evaluative ones. In the standard market setting of microeconomic theory, for example, the analyst may offer a model for explanatory, design, and evaluative purposes. Indeed, in this context the descriptive, explanatory, and normative conceptions of preference tend, or perhaps appear, to coincide. In the standard model, consumers have preferences over (their own) consumption bundles. In this context, it is plausible to believe that the agent's preferences both explain and summarize her choices. Moreover, it is plausible to be-

51 See George L. Priest, The Common Law Process and the Selection of Efficient Rules, 6 J. LEGAL STUD. 65 (1977) (claiming that enforceable legal rules are more often efficient due to their unique endurance and response to administrative pressures); Paul H. Rubin, Why Is the Common Law Efficient?, 6 J. LEGAL STUD. 51, 51 (1977) (arguing that efficient rules emerged through an "evolutionary mechanism" controlled by disputants' utility-maximizing decisions).

52 See Robert Cooter & Lewis Kornhauser, Can Litigation Improve the Law Without the Help of Judges?, 9 J. LEGAL STUD. 139, 156 (1980) ("We have not found a compelling reason to think that inefficient laws will be litigated more frequently . . . .").

53 Id. at 147-50 (suggesting that differential litigation will select efficient rules only under special conditions); see also Lewis A. Kornhauser, Notes on the Logic of Legal Change, in SOCIAL RULES: ORIGIN; CHARACTER; LOGIC; CHANGE 169, 173-77 (David Braybrooke ed., 1996) (illustrating that the conjecture that differential litigation leads to the selection of efficient rules remains partially open). Gillian K. Hadfield argues that efficient rules are not selected even if judges seek efficiency because the set of litigated cases is a biased sample of the set of conducts governed by a rule. Gillian K. Hadfield, Bias in the Evolution of Legal Rules, 80 GEO. L.J. 583 (1992).

54 When the choice situations are sufficiently simple, explanatory preferences will correspond to descriptive preferences because the agent's motivations will directly determine the agent's choices.

55 The model thus places a restriction on the structure of each agent's preference; it forbids "altruism" or concern about one's relatives and friends.
lieve that, in the restricted world of the model, satisfaction of the agent's preferences constitutes her well-being.56

A typical model in economic analysis of law seeks not only to explain behavior but also to propose improvements in the legal institution under study. The same economic model of accident law, for instance, serves both to explain the pattern of decisions that individuals make concerning the choice of levels of care and activity in risky situations and to evaluate different rules of tort law that govern these accident situations. This neat coincidence of summary, explanation, and evaluation need not always hold.57 Such a failure of the three sets of preferences to coincide presents both complications and an opportunity to explain the anomalies of preference theory that behavioral economics has documented.

Consideration of the formalism underlying preference identifies two fundamental ways in which the coincidence of preference may fail. Consider, for example, the relation between explanatory and evaluative preferences, with explanatory preferences as the pair \((D_e, R_e)\) and evaluative preferences as the pair \((D_w, R_w)\) where \(W\) signifies well-being. Explanatory preferences might differ from evaluative preferences because the domains differ; i.e., \(D_e \neq D_w\). Alternatively, though the domains coincide, the relations \(R_e\) and \(R_w\) may differ. The relation between the domain of preference and the domain of choice raises additional issues. To begin, one might ask what function the descriptive conception of preference as a summary of choice serves. Paul Samuelson clearly had an explanatory (as well as design and evaluative) aim in mind,58 but the descriptive conception of choice cannot explain non-tautologically why a particular agent chose as she did. Rather, the descriptive conception of preference aims to explain (or reveal) the interdependence among choices; that is, it hopes to explain not individual choice but market outcomes.

56 In fact, some models of consumer behavior and product differentiation do not assume that the agent has preferences directly over consumption bundles or commodities. Rather, these models assume that each agent has preferences over attributes of products. Preferences over products (if they exist) are thus derivative of these more primitive preferences. See Kelvin Lancaster, Consumer Demand 140 (1971) (explaining that certain characteristics of goods are "relevant" in that they affect "predictions about the choice or ordering of goods by consumers").

57 In fact, in this context the interpretation of preference will also coincide with the normative project of characterizing rational behavior.

B. Distinguishing the Domain of Choice from the Domain of Preference

Applications of preference theories require a concrete interpretation of the domain $D$ of preference. Earlier I noted that, in the standard context of application, concrete interpretations of preference have a common feature: they identify the domain of preference with the domain of choice (and with the domain of evaluation). In standard applications, the set of options from which agents must choose—the domain of choice—is a subset of the set of objects over which the agent has (explanatory) preferences. This domain of preference in turn corresponds to the domain of evaluation.

Many choice situations, however, do not have this structure. Often the domain of options over which an agent chooses is not a subset of the domain of elements over which she has fundamental preferences. Similarly, an explanation of the agent's choice may not correspond to the elements of the situation that contribute to or constitute the agent's well-being.

Consider, for example, six friends at an Indian restaurant. They have agreed to share food but recognize that group deliberation over the contents of the meal will be both long and controversial. Therefore, they agree that each of them will choose one course. In this situation, a reasonable model would assume that each friend has a preference over full meals—i.e., over combinations of six dishes. Yet each friend chooses only one dish, or one element of that meal.

Consider a second example illustrating a different relation that might hold between the domains of choice and preference. In many countries, students must choose to specialize their education at a young age. A student at age twelve, for example, might have to choose to concentrate her studies on mathematics and science rather than languages and literature. This choice is difficult for many reasons. The individual—call her Liza—may be unsure about many aspects of the future, including her own talents and the nature of the labor market she will enter. Even taking this uncertainty into account, however, the domain of choice still differs from the domain of preference.

One may assume, however implausibly, that, at age twelve, Liza has preferences over the careers that she might pursue. At twelve, however, she must choose not a career but rather an educational program. Each educational program leads to a different set of possible careers. If she has preferences over educational programs, these preferences must derive in some way from her preferences over careers.
One might think this discrepancy between the domains of choice and of preference is unimportant because the derivation of preferences over educational programs from preferences over careers is straightforward. Liza prefers educational program $A$ to educational program $B$ if and only if she prefers the career $C(A)$ to the career $C(B)$, where $C(A)$ is the career that Liza most prefers among those available to one who pursues program $A$ and $C(B)$ is the career that Liza most prefers among those available to one who pursues program $B$. This method of ranking sets of careers has the most plausibility in a world in which Liza will have equal and certain access to each of the careers "made available" by an educational program. Liza more likely faces a world in which her success in an educational program is uncertain and in which not all careers that possibly flow from an educational program will be equally available to her. In this world, Liza might not evaluate educational programs $A$ and $B$ in terms of the best (or most desirable) career $C(A)$ and $C(B)$ that the program opens for her.\footnote{She might compare program $A$ to program $B$ in terms of the value of the "expected career" to which each program leads, with each career weighted by her beliefs about the likelihood she would follow that career and by her beliefs concerning the extent to which she would find that career fulfilling. It remains the case, however, that even under this decision criterion, her choice of educational programs differs from the domain over which she has preferences.}

The domain of choice $D_e$ then may have at least two relations to the domain of explanation $D_e$. In the first example, where a group of individuals chose a meal, each element of $D_e$ was one point in a multidimensional space. The object over which the agent had preferences (meals) had distinct features (each of the six courses of the meal). Each individual chose a feature of the meal, not the meal itself. The domain of choice is thus, in some sense, smaller than the domain of preference. In the second example, Liza had (primitive) preferences over careers, but in choosing educational programs, she chose not a career but a set of possible careers. In this instance, then, the domain of choice consisted of larger objects than those over which she had primitive preferences.

An important class of models in economics has the structure of the meal example, in which the domain of choice consists of features of the object over which the agent has (primitive) preferences. Consider an individual Henry who will live for $T$ periods. He must determine his consumption in each period $t$. The standard economic model assumes that Henry has preferences over consumption paths;
i.e., he cares about his consumption over his lifetime. \( D_e \) thus consists of all possible vectors \( c = (c_0, c_1, \ldots, c_T) \). Henry, however, does not generally choose his lifetime consumption pattern \( c \) at once; rather, in each period \( t \), Henry chooses \( c_t \).

Behavioral economics has studied some members of this class of models because they pose particular difficulties for preference theorists. Often, for example, behavioral economists point to models of this type to illustrate a divergence between “rational” behavior and how individuals actually behave.\(^60\) The fact that individuals do not conform to criteria of rationality may undermine preference theory as an explanatory theory, but it need not undermine it as a normative account of rationality.\(^61\)

A special feature of these consumption models is that, over time, they conceal some complexities that may arise when the domain of choice consists of features of elements of the domain of explanation. In particular, other examples indicate how this structure may provide additional resources to preference theories to account for the anomalies that behavioral economics has exposed.\(^62\)

C. An Example

In this section I employ an example used in my previous work for two purposes.\(^63\) First, I illustrate how distinguishing the domain of

---

\(^60\) This divergence does not emerge solely from the structure of preferences already specified. It emerges, rather, from the way in which consumption in one period is related to consumption in another period. If individuals do not discount the future in the appropriate way, then they confront a problem of “time inconsistency.” The classic work on time inconsistency is R.H. Strotz, *Myopia and Inconsistency in Dynamic Utility Maximization*, 23 REV. ECON. STUD. 165 (1956). For a recent empirical and theoretical study, see George Ainslie, *Breakdown of Will* (2001). A problem of time inconsistency arises when the consumption plan that the agent finds optimal at time \( t \) is no longer optimal when she reaches time \( t+1 \) (or later time).

\(^61\) More precisely, the evidence of nonconformity of individual behavior to the normative requirements of preference theories may undermine only preference theory as a psychological explanation of behavior; it does not directly undercut nonpsychological explanatory accounts.

\(^62\) Some psychologists studying sequential choice have noted that the studies of time consistency assume that the agent’s preference over the stream of consumption is separable across time. George F. Loewenstein & Dražen Prelec, *Preferences for Sequences of Outcomes*, 100 PSYCHOL. REV. 91, 91-108 (1993).

choice from the domain of explanation might explain apparent intransitivities if the domain of explanation is identified with the domain of choice. Second, the example suggests how distinguishing the domain of choice from the domain of explanation may provide new explanatory insights and point to novel approaches to the design of institutions.

Consider an election of an assembly—a committee, a city council, a state legislature, a national parliament—to govern some body or jurisdiction. Usually election procedures provide that a voter casts a ballot for a candidate (or candidates) and that the assembly is filled candidate by candidate. In the simplest context, to which I will generally refer, the assembly is elected at large—i.e., each member of the electorate casts a vote for each “seat” in the assembly (and the candidates need not declare for a particular seat).

Most models of elections consider the usual election procedure. The models generally assume that each voter has preferences over candidates. That is, the models assume that each agent’s domain of choice—the set of candidates for whom she may vote—is identical to the agent’s domain of preference. Reflection, however, suggests that this “natural” assumption is not obviously appropriate if one wishes either to explain (psychologically at least) or to evaluate these election procedures.

A voter’s well-being does not depend directly on the candidate who is elected. The voter’s well-being will most likely depend directly on the legislative program enacted by the assembly. This suggests that the voter’s basic (evaluative) preferences should be over legislative programs or over assemblies as a whole, not candidates.

---

issue of voting rights from an assembly-preference perspective, and arguing that the traditional focus on candidate-preference is misleading. I draw here on the later text. For other work that adopts the same approach to voting, see Jean-Pierre Benoit & Lewis A. Kornhauser, Social Choice in a Representative Democracy, 88 AM. POL. SCI. REV. 185 (1994) [hereinafter Benoit & Kornhauser, Social Choice]; Jean-Pierre Benoit & Lewis A. Kornhauser, On the Separability of Assembly Preferences, 16 SOC. CHOICE & WELFARE 429 (1999) [hereinafter Benoit & Kornhauser, Separability].

64 A similar analysis applies to districted and numbered-seat elections in which candidates run for specified seats and in which, in the case of districts, the electorate for each district may differ. In addition, the model applies to systems of party-list proportional representation in which a voter casts a ballot for a party. The number of votes for the party determines the number of seats it receives in the assembly. In this context, each party is a candidate.

65 If the voter is the candidate, or related to the candidate, she might derive direct benefits from the election of the candidate. Alternatively, the election of a particular candidate would have some intrinsic, expressive, or symbolic value. Neither possibility accounts for the interests of most voters.
The above argument extends to explanatory preferences as well. A self-interested voter is unlikely to care about which candidate is elected. Rather, she will care about what the assembly does for her. Again, this entails that basic explanatory preferences should be over the domain of either assemblies or legislative programs.66

Though I assume throughout that voters cast ballots nonstrategically,67 I must clarify the nature of nonstrategic voting. Such clarification requires a distinction between two ways in which the agent may vote nonstrategically (or sincerely) for a legislative program.68 She might cast her ballot as if her vote were decisive. Alternatively, she might cast her ballot for the legislative program that she most prefers. When individuals vote for assemblies or legislative programs directly, these two accounts of sincere voting coincide. When the individual votes for candidates, however, the two justifications for sincere voting may differ. I shall say that, when an individual votes for the candidate that supports the legislative program she most prefers, the individual votes expressively. The voter's preferences over legislative programs induce an expressive ranking of candidates. When an individual votes for candidates as if her ballot were decisive, she may not be able to rank candidates unequivocally.69 When she can rank them in this way, I

---

66 Explanatory or evaluative preferences might range over assemblies rather than legislative programs for at least two different reasons. First, the voter might desire a representative assembly. She might have this preference because she cares about fair processes and thinks that a representative assembly is fair. Second, the voter may not know what issues will confront the assembly. She may face radical uncertainty concerning the legislative program that will be enacted. The uncertainty is radical because she may not even know the state space, i.e., she may be ignorant of the set of possible issues that the assembly will confront.

67 This assumption may be defended in a number of ways. First, most voting games have many equilibria even if there are only two candidates. In a voting game over two candidates, there may be many equilibria in which a voter may, in equilibrium, vote for her less preferred candidate. Second, when electorates are large, the calculation demands on the agent are very large, and it is implausible to believe that she can solve the strategic problem.

68 In contexts in which the legislative program facing the assembly is fixed, there is a natural correspondence between legislative programs and assemblies.

69 Benoit and Kornhauser use an example to show that separable assembly preferences—the condition for the existence of well-defined simple preferences over candidates—are rare. See Benoit & Kornhauser, Social Choice, supra note 63, at 186 (arguing that many separable assembly rankings are actually "consistent with any single candidate ranking" and are, therefore, not unique). They further show that such preferences do not generally derive from the usual assumptions on preferences over legislative programs. See Benoit & Kornhauser, Separability, supra note 63, at 437 (concluding that the assumption of separable assembly preferences is not warranted "if assembly preferences are derived from preferences over legislative outcomes").
shall say that the individual votes simply and that the derived preferences over candidates yield her simple ranking.

I assume that voters care about legislative programs and that these preferences induce preferences over the various legislatures. Specifically, suppose that the legislature will act on four issues: employment discrimination, environmental regulation, economic stimulus, and civil liberties. Assume further (and without loss of generality) that, on each substantive issue, the legislature can adopt only one of two possible positions: a liberal one $L$ or a conservative one $C$. Consequently, there are sixteen possible legislative programs, each one defined by the choice of position $L$ or $C$ on each of the four issues. The most liberal program then is $(L, L, L, L)$ and the most conservative $(C, C, C, C)$. Finally, assume that each voter's preferences are roughly "spatial"—i.e., she has a most preferred program $p^*$, and she prefers a program $p$ to a program $q$ if and only if $p$ is "closer" to $p^*$ than $q$ is to $p^*$. The voter breaks ties lexically. 

1. Voters May Have Multiple Derived Preferences

Consider a jurisdiction holding at-large elections for a three-person assembly. Each voter may cast a ballot for only one candidate. Each candidate can be identified with the legislative position that she adopts. The following four candidates contest the election:

- $w$: $(L, L, L, C)$
- $x$: $(L, L, C, L)$
- $y$: $(C, C, L, L)$
- $z$: $(L, L, C, C)$

With these four candidates, only four (of the sixteen possible) assemblies can be elected:

- $A = \{w, x, y\} = (L, L, L, L)$
- $B = \{w, y, z\} = (L, L, L, C)$
- $C = \{x, y, z\} = (L, L, C, L)$
- $D = \{w, x, z\} = (L, L, C, C)$

---

70 Suppose Liza is a liberal voter; she most desires a legislative program $(L, L, L, L)$. She prefers any legislative program that adopts more liberal positions to one which adopts fewer. For example, she prefers the program $(L, L, L, C)$ to the program $(L, L, C, C)$. Beyond this, she cares marginally more about the first issue than the second, the second more than the third, and the third more than the fourth, in such a way that, if two programs have the same number of liberal positions, she prefers the one whose first $C$ comes latest. For example, she prefers the program $(L, C, C, L)$ to $(C, L, L, C)$.

71 Assume that each candidate, if elected, will adhere to her announced position in her legislative votes.
Consider a liberal voter Liza. She ranks the assemblies $A > B > C > D$, with $A$ as her most preferred assembly. In an assembly-based procedure, a vote for $A$ would be expressive of the liberal’s view and sincere in the economist’s usual sense. It would be simple in our sense as well, though trivially, since Liza could not regret the election of the assembly $A$.

However, the election is candidate-based, and Liza can cast only one vote. It may appear that she should vote for candidate $w$ who holds the views closest to her own (taking into account the order of importance of the issues). Indeed, her expressive ranking of the candidates is $w > x > z > y$. This ranking is not her simple ranking as Liza would prefer to complete the assembly $\{x, z\}$ with candidate $y$ rather than with candidate $w$. In fact, her simple ranking of the candidates is $y > w > x > z$.\(^2\) The candidate ranked lowest expressively is ranked highest simply. This simple ranking, rather than the expressive ranking, is the “correct” consequentialist ranking because Liza will never regret a decisive vote that gets $y$ elected, though she might regret a decisive vote for $w$. For instance, if she votes for $w$ and hence the assembly $\{w, x, z\}$ is elected, she will regret not having voted for $y$ if that vote would have resulted in the assembly $\{x, y, z\}$ instead.

The reason that Liza simply ranks $y$ first, even though $y$ is expressively furthest from her, is clear. She dislikes $y$ because of $y$’s views on the first two issues. However, $y$ is “extreme” on these two issues—all the other candidates disagree with $y$ on these issues and will always vote $y$ down. On the “swing” issues, $y$ agrees with our voter. With many issues to be voted upon, a voter could favor a candidate who is quite far from her ideal position.\(^3\)

Finally, suppose that the voter could cast votes for three candidates. If she voted expressively, she would cast her ballot for candidates $w$, $x$, and $z$. Should these three candidates prevail, she would surely regret her vote because the assembly $\{w, x, z\}$ is the one she least prefers. Had she voted simply for $w$, $x$, and $y$, she would not regret the success of any of the candidates for whom she voted, and she would be delighted if all three prevailed as she most prefers the assembly $\{w, x, y\}$.

---

\(^2\) Although the simple and expressive rankings differ here, this is not always the case. Also, in this example, the simple vote also proves to be a dominant strategy. This coincidence between simplicity and dominance is not a general feature of simplicity.

\(^3\) In fact, a “crackpot” candidate could gather significant support precisely because many of her proposals will never be enacted; only her views on close issues are germane.
Notice that the expressive and simple rankings correspond to distinct conceptions of rationality. One might thus connect them to normative as well as explanatory preferences. The simple ranking of candidates corresponds to a consequentialist conception of rationality in which the individual votes for the candidate that most promotes her aims. The expressive ranking, by contrast, corresponds (at least on some interpretations) to an expressive conception of rationality. As the prior paragraph indicates, when the individual votes expressively, she votes contrary to her consequentialist interests.

2. Individuals May Rationally Violate Conditions on Preferences over Candidates

Two extensions of this example suggest how one might observe "inconsistent choices"—i.e., choices that arguably violate transitivity, completeness, asymmetry, or the weak axiom of revealed preference—even though the individual acts optimally. In both extensions, I assume that the individual votes consequentially rather than expressively. The first extension keeps the set of candidates fixed but changes the size of the assembly; the second extension keeps the size of the assembly fixed but alters the set of candidates. Each change induces a change in the individual's voting behavior.

a. First extension

Suppose that the jurisdiction has a two-person assembly rather than a three-person assembly. The assembly functions by majority rule; in the event of ties, the status quo \((C, C, C, C)\) prevails. As before, only \(w, x, y,\) and \(z\) are contesting the election. There are now six possible assemblies with the associated legislative programs:

\[
A' = \{w, x\} = (L, L, C, C) \\
B' = \{w, y\} = (C, C, L, C) \\
C' = \{w, z\} = (L, L, C, C) \\
D' = \{x, y\} = (C, C, C, L) \\
E' = \{x, z\} = (L, L, C, C)
\]

\[ F' = \{y, z\} = (C, C, C, C) \]

Liza ranks these assemblies as follows: \( A' = C' = E' > B' > D' > F' \). Liza would never regret a decisive vote for \( w \).

If the elected assembly consists of only two people, Liza’s simple ranking of candidates is \( w > x > z > y \). If the individual can vote for only one candidate, she would choose \( w \) from the set \( \{w, x, y, z\} \). Recall, however, that, with a three-person assembly, the voter’s simple ranking is \( y > w > x > z \). In a three-person election in which the individual can vote for only one candidate, she would choose \( y \). This pair of choices is “irrational” on the assumption that the individual has basic preferences over candidates.

Of course, one might argue that the agent acts rationally by arguing that the two choice situations are not equivalent. In one, the agent votes for a two-person assembly while in the other she votes for a three-person assembly. Phrased differently, one might say that Liza has basic preferences not over candidates simpliciter, but candidates given the slate contending the election. This observation is true, but it points to the superiority of a model that derives preferences over candidates from basic preferences over legislative programs, as this model explains why preferences over candidates have changed rather than making an ad hoc assumption. Indeed, the attempt to redescribe the domain of preferences as “candidates within a slate” effectively admits that the domain of preferences differs from the domain of choice.

b. Second extension

In the examples thus far, the voter has well-defined preferences over candidates. The existence of these well-defined preferences depends on the number (and identity) of candidates. Suppose a fifth candidate \( s \) contests the three seats in the at-large assembly already contested by candidates \( w, x, y, \) and \( z \). Let this candidate \( s \) endorse the legislative program as follows:

\[ s = (L, C, L, L) \]

These five candidates may constitute ten different three-person assemblies (different in personnel, not necessarily in legislative program) including \( A, B, C, \) and \( D \). Consider two other assemblies that

---

25 The additional six assemblies are:

\[ E = \{s, w, x\} = (L, L, L, L) \]
\[ F = \{s, w, y\} = (L, C, L, L) \]
\[ G = \{s, w, z\} = (L, L, L, C) \]
are possible:

\[ E = \{s, w, x\} = (L, L, L, L) \]
\[ F = \{s, w, y\} = (L, C, L, L) \]

Liza with ideal point \((L, L, L, L)\) can no longer unambiguously rank candidates. She prefers \(x\) to \(y\) if \(s\) and \(w\) are elected (because she prefers assembly \(E\) to assembly \(F\)), but she prefers \(y\) to \(x\) if \(w\) and \(z\) are elected (because she prefers assembly \(B\) to assembly \(D\)). Suppose she believes that \(s\) and \(w\) will be elected; then she best promotes her legislative program by voting for \(x\). So out of the set \(\{s, w, x, y, z\}\), she chooses \(x\). Previously, however, I showed that, from the set \(\{w, x, y, z\}\), she would vote for \(y\). This pair of choices violates the weak axiom of revealed preference because it violates contraction consistency. Nonetheless, Liza's choices conform to preference theory when the domain of choice is appropriately defined.

**CONCLUDING REMARKS**

For several reasons, the concept of preference cannot be easily dislodged from the central role it plays in economic theory. First, the concept furthers a number of distinct, but interrelated, projects. The normative uses of the concept of preference as measures of well-being and as criteria of rational decision making support the use of the concept for explanatory purposes. Substitution of psychological explanations that do not rely on the concept of preference severs the links between these projects. Second, apparent violations of the conditions that the concept of preference imposes on choice may dissolve once one recognizes that the domain of choice need not correspond to the domain of preference. Consequently, one must interpret the experimental findings with care. Preference theory may prove an adequate psychological theory once we correctly specify the domain of preference.

\[ H = \{s, x, y\} = (L, C, L, L) \]
\[ I = \{s, x, z\} = (L, L, L, C) \]
\[ J = \{s, y, z\} = (L, C, L, L) \]

The assemblies \(F = H = J\) yield the only new legislative program.

\(^{76}\) Similarly, she prefers \(w\) to \(y\) if \(s\) and \(x\) are elected (because she prefers \(E\) to \(H\), which has the same legislative program as \(F\)) and \(y\) to \(w\) if \(x\) and \(z\) are elected (because she prefers \(C\) to \(D\)).

\(^{77}\) *Supra* note 19.